

09/868760

Attorney Docket No. 574-2001

New U.S. Patent Application filed June 21, 2001 for application entitled: **531 Rec'd PCT 21 JUN 2001**

SERINE PROTEASE INHIBITORS

corresponding to PCT/NZ99/00227

filed December 23, 1999

Express Mail No.: EL742691527US

Date of Deposit: June 21, 2001

I hereby certify that this application and the accompanying papers are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to:

Box PCT
Assistant Commissioner for Patents
Washington, D.C. 20231

Charles Jackson

SERINE PROTEASE INHIBITOR

This invention relates to a protein and compositions which contain it. More particularly, it relates to a protein which *inter alia* exhibits activity as a metal cation
5 binding agent and as an anti-thrombin agent.

BACKGROUND

Thrombin is a serine protease involved in blood coagulation. It has specificity for the
10 cleavage of arginine-lysine bonds as well as cleaving an arginine-threonine bond in pro-thrombin, releasing pre-thrombin which is subsequently cleaved to produce active thrombin. This active thrombin can then release more thrombin from pro-thrombin. In blood clotting and coagulation, thrombin cleaves fibrinopeptide B from fibrinogen as well as converting blood factors IX to IXa, V to Va, VIII to VIIIa and XIII
15 to XIIIa.

Inhibitors of thrombin therefore inhibit coagulation and have application in any procedure where coagulation is undesirable. One such application is in the collection and storage of blood products. Another is in medicaments for preventing
20 or reducing coagulation for example in treating or preventing cardiac malfunctions.

Anti-thrombin agents are known. One example is anti-thrombin III (AT-III). However, AT-III is capable of effectively inhibiting thrombin only in the presence of
25 heparin.

The applicants have now identified a novel protein which has a range of activities, including anti-thrombin activity, and which when active against thrombin does not require heparin as a cofactor. It is towards this protein that the present invention is
30 broadly directed.

SUMMARY OF THE INVENTION

In a first aspect, the present invention provides an isolated protein which has a molecular weight of about 55 kDa and an amino acid sequence which includes one or more of the following:

- (a) DGEQCNDGQN (SEQ ID NO. 1)
 (b) QGGHEVESERVACCVIGRA (SEQ ID NO. 2)
 (c) GQSHPEIVH (SEQ ID NO. 3)
 (d) YHGHDDA (SEQ ID NO. 4)
 5 (e) VVNEVHH (SEQ ID NO. 5),

or an active fragment thereof.

In a further aspect, the invention provides an isolated protein which comprises the amino acid sequence of

10 D G E Q C N D G Q N K D D H H D D H H D D H H D D H D D D D
 E T M H Y A Q C E M E P N P H M A S S L H H H V H G S I E L
 S Q K G H G A V Y L E L H L V G F N T S E D H D D H H H G L
 H L H M L G D M S A G C D S I G E L Y N A H P E K H A D P G
 D L G D L V D D D R G V V N E V H H Y A W L D I D G T A P N
 15 T E A L I G H S M T I L Q G S H T D A D T P A S R I A C C V
 I G H G K A R P E T A A A L H H E L E E D K T E H Y A H C D
 V R S N T H Q P K A L H H H V H G T I D F K Q V G Y G D L E
 V S Y H L E G F N V S D D H K D H L H D V Q I Y A N G D L T
 S G C D N L G A K Y D P H E D Y H S E L G D L G D I H D D D
 20 H G V V N E S H R Y S W I N I F G D D S V L G R S I A I H Q
 R D H L H K S A K I A C C V I G R G Q S H P E I V H R A K C
 V V R P N T E S T G L H H H V S G S I T F E Q T P G G S T H
 M T A D L K G F N V S E D L S H H R H G V Q L H E W G D M S
 H G C H S L G R M Y H G H D D A H D P K R P G D L G D V I D
 25 D S H G I V H S T R T F D H L N V E D L N A R S L V I M Q G
 G H E V E S E R V A C C V I G R A (SEQ ID NO. 7)

or an active fragment thereof.

In yet a further aspect, the invention provides an isolated protein which is obtainable from the haemolymph of *Perna canaliculus* which has an apparent
 30 molecular weight of 75 kda determined by PAGE, or an active fragment thereof.

Conveniently, said protein or fragment has activity as:

- (i) a serine protease inhibitor; or

(ii) a divalent cation binding agent.

The invention further provides a protein which is a functionally equivalent variant of a protein or fragment as defined above.

Still further, the invention provides a protein which is obtainable from a shellfish
5 other than *Perna canaliculus* and which is a functionally equivalent variant of a protein or fragment as defined above.

In another aspect, the invention provides a polynucleotide encoding a protein or fragment as defined above.

The polynucleotide may comprise the nucleotide sequence of

10

15

20

25

30

5' GAYGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGA
CCACCACGATGATCACCATGACGACCATGATGATGATGATGAAACAATGCACT
ATGCCCAGTGTGAAATGGAACCAACCCTCATATGGCTAGCAGCCTTCACCA
CCATGTCCATGGCAGCATAGAGTTGTACAGAAGGGTCATGGAGCTGTTTAT
CTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA
TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTA
TTGGCGAACTGTACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCT
CGGTGACCTGGTTGACGATGATAGGGGCGTGGTTAATGAAGTTCATCATTATG
CTTGTTGGACATTGATGGTACAGCACCAACACCGAAGCTCTCATTGGACA
CTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCCAGCCAGTA
GAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGC
TGCTCTACATCACGAGCTAGAGGAAGATAAACTGAGCATTATGCCCATTGTG
ACGTAAGATCTAATACACACCAACCAAAAGGCTCTTCATCATCATGTCCACGGA
ACCATCGATTTCAAACAAGTTGGTTATGGTGACCTTGAAGTGTCTACCATTTA
GAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGACGTACAGAT
CTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATAT
GATCCTCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGA
TGATGACCATGGCGTTGTCAATGAAAGCCACAGATATTCCTGGATCAATATCT
TCGGTGATGACAGTGTCTGGGACGTTCTATTGCCATTACCAAAAGAGACCAT
CTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACGTGGACAGAGCCA
TCCAGAAATTGTTACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTAC
TGGTTTACATCACCATGTCTCTGGTTCTATAACATTCTGAACAGACCCCTGGAG

GATCAACACATATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTG
 TCACATCATCGTCATGGTGTGCAGCTCCATGAATGGGGAGATATGTCCCATG
 GCTGTCACTCCTTAGGCAGAATGTACCATGGTCATGATGATGCTCATGACCCC
 AAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATCGTTCA
 5 TTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCT
 TGTGATTATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGT
 GTTATAGGACGGGCA (SEQ ID NO. 6)

or a variant thereof.

Still further, the invention provides a vector or construct which includes a
 10 polynucleotide as defined above.

In another aspect, the invention provides a composition which comprises a protein
 or fragment as defined above.

The composition may be a medicament, a food, a dietary supplement, (optionally
 including the protein associated with or bound to at least one divalent cation of
 15 dietary significance) or a bioremediation agent.

In still another aspect, the invention provides a process for obtaining a protein as
 defined above which comprises the step of centrifuging material containing *Perna*
canaliculus haemolymph or an extract thereof and recovering the sedimented
 protein.

20

DESCRIPTION OF THE DRAWINGS

While the present invention is broadly as defined above, it also includes
 embodiments of which the following description provides examples. In particular, a
 25 better understanding of the present invention will be gained through reference to
 the accompanying drawings in which

Figure 1: Purification of pernin from mussel haemolymph

30 a) light-scattering band following centrifugation of *P. canaliculus* haemolymph
 in CsCl; haemolymph was first centrifuged at low speed to remove

haemocytes and then at high speed; the re-suspended pellet was then centrifuged in CsCl.

5 **b)** UV absorption profile (254 nm wavelength) from fractionation of the CsCl gradient; the light-scattering material in figure 1a appears as a peak.

10 **c)** protein composition in 1 ml fractions of a CsCl gradient following electrophoresis in a 12% polyacrylamide gel; the heavily stained (Coomassie) bands coincide with the position of the light-scattering and UV-absorbing regions of the gradient; the molecular weight was approximately 75 kDa as compared with polypeptide molecular weight standards (lane 6) (refer Figure 4a for standards). Lanes 1-5 and 7-9 contained samples from the CsCl gradient.

15 **Figure 2:** Virus-like particles observed by transmission electron microscopy of material in light scattering band in a CsCl gradient. Bar in micrograph represents 100 nm.

20 **Figure 3:** HPLC elution profile of pernin at 280 nm wavelength purified by CsCl gradient centrifugation..

Figure 4: SDS-PAGE profiles (12% gels) of aggregating protein species from *P. canaliculus* and other shellfish species

25 **a)** proteins extracted from whole shellfish and purified as described in Materials and Methods: lane 1: molecular weight standards (Bio-Rad, USA) :**pb** phosphorylase B, 97.4 kDa; **bsa** bovine serum albumin, 66 kDa; **ova** ovalbumin, 45 kDa; **ca** carbonic anhydrase, 31 kDa; lane 2: Greenshell™ mussel *P. canaliculus*; lane 3: blue mussel *Mytilis edulis*; lane 4: oyster *Crassostrea gigas*; lane 5: pipis *Paphies australis*.

30

b) PAGE analysis of human transferrin (Sigma, USA, MW ca. 80 kDa), a glycosylated protein, and pernin from *P. canaliculus* following treatment with endoglycosidase-F: lane 1: untreated transferrin; lane 2: transferrin treated

with glycosidase-F; lane 3: untreated pernin lane 4: pernin treated with glycosidase-F.

Figure 5: Activity of *P. canaliculus* haemolymph protein following centrifugation in a 30 kDa molecular weight exclusion filter for 10 min at 1000 *g* (Ultrafree-MC filter, 30,000 MW exclusion, Millipore, USA)

a) SDS-PAGE profile of haemolymph protein at various stages of purification. Lane 1: "crude" haemolymph (haemocytes removed); lane 2: resuspended pellet after ultracentrifugation of "crude" haemolymph for 80 min at 250,000 *g*; lane 3: pernin retentate; lane 4: filtrate (no proteins evident); lane 5: molecular weight markers, (refer Figure 4a); lanes 6,7: 10-fold dilutions of samples from lanes 2 and 3.

b) Anti-thrombin activity of 30,000 MW exclusion filter retentate and filtrate.

con⁺ = the standard 1/41 dilution of human plasma (i.e. standard anti-thrombin III activity);

con⁻ = thrombin with no added plasma (buffer control); **filtrate:** material passed through a 30,000 MW exclusion filter;

retentate: pernin protein retained by exclusion filter.

DESCRIPTION OF THE INVENTION

As broadly outlined above, in one aspect the present invention provides a novel protein. The protein of the invention has an apparent molecular weight of 75 kDa, calculated by polyacrylamide gel electrophoresis (PAGE). The molecular weight inferred from the gene sequence is approximately 55 kDa.

One specific protein of the invention was initially identified as an extract from the New Zealand green lipped mussel *P. canaliculus*. It is therefore obtainable by extraction directly from *P. canaliculus*.

This protein has the amino acid sequence of SEQ ID NO. 7.

The protein of the invention can include its entire native amino acid sequence or can include only parts of that sequence where such parts constitute fragments which remain biologically active (active fragments). Such activity will normally be as a serine protease inhibitor or a divalent cation binding agent but is not restricted to these activities.

The invention also includes within its scope functionally equivalent variants of the protein of SEQ ID NO. 7.

10 The phrase "functionally equivalent variants" recognises that it is possible to vary the amino acid of a protein while retaining substantially equivalent functionality. For example, a protein can be considered a functional equivalent of another protein for a specific function if the equivalent peptide is immunologically cross-reactive with and has at least substantially the same function as the original protein.

15 The functionally equivalent protein need not be the same size as the original. The equivalent can be, for example, a fragment of the protein, a fusion of the protein with another protein or carrier, or a fusion of a fragment with additional amino acids. It is also possible to substitute amino acids in a sequence with equivalent amino acids using conventional techniques. Groups of amino acids normally held to be equivalent are:

- (a) Ala, Ser, Thr, Pro, Gly;
- (b) Asn, Asp, Glu, Gln;
- 25 (c) His, Arg, Lys;
- (d) Met, Leu, Ile, Val; and
- (e) Phe, Tyr, Trp.

Polypeptide sequences may be aligned, and percentage of identical amino acids in a specified region may be determined against another sequence, using computer algorithms that are publicly available. The similarity of polypeptide sequences may be examined using the BLASTP algorithm. BLASTP software is available on the NCBI anonymous FTP server (ftp://ncbi.nlm.nih.gov) under /blast/executables/. The use of the BLAST family of algorithms, including BLASTP, is described at NCBI's website at URL <http://www.ncbi.nlm.nih.gov/BLAST/newblast.html> and in the publication

of Altschul, Stephen F., *et al.* (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-34023.

The protein of the invention together with its active fragments and other variants
5 may be generated by synthetic or recombinant means. Synthetic polypeptides
having fewer than about 100 amino acids, and generally fewer than about 50 amino
acids, may be generated by techniques well known to those of ordinary skill in the
art. For example, such peptides may be synthesised using any of the commercially
10 available solid-phase techniques such as the Merryfield solid phase synthesis
method, where amino acids are sequentially added to a growing amino acid chain
(see Merryfield, J. Am. Chem. Soc 85: 2146-2149 (1963)). Equipment for automative
synthesis of peptides is commercially available from suppliers such as Perkin
Elmer/Applied Biosystems, Inc. and may be operated according to the
manufacturers instructions.

15 The protein, or a fragment or variant thereof, may also be produced recombinantly
by inserting a polynucleotide (usually DNA) sequence that encodes the protein into
an expression vector and expressing the protein in an appropriate host. Any of a
variety of expression vectors known to those of ordinary skill in the art may be
20 employed. Expression may be achieved in any appropriate host cell that has been
transformed or transfected with an expression vector containing a DNA molecule
which encodes the recombinant protein. Suitable host cells includes procaryotes,
yeasts and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*,
yeasts or a mammalian cell line such as COS or CHO, or an insect cell line, such as
25 SF9, using a baculovirus expression vector. The DNA sequence expressed in this
matter may encode the naturally occurring protein, fragments of the naturally
occurring protein or variants thereof.

DNA sequences encoding the protein or fragments may be obtained by screening an
30 appropriate *P. canaliculus* cDNA or genomic DNA library for DNA sequences that
hybridise to degenerate oligonucleotides derived from partial amino acid sequences
of the protein. Suitable degenerate oligonucleotides may be designed and
synthesised by standard techniques and the screen may be performed as described,
for example, in Maniatis *et al.* Molecular Cloning - A Laboratory Manual, Cold
35 Spring Harbour Laboratories, Cold Spring Harbour, NY (1989). The polymerase

chain reaction (PCR) may be employed to isolate a nucleic acid probe from genomic DNA, a cDNA or genomic DNA library. The library screen may then be performed using the isolated probe.

- 5 Variants of the protein may be prepared using standard mutagenesis techniques such as oligonucleotide-directed site specific mutagenesis.

A specific polynucleotide of the invention has the nucleotide sequence of SEQ ID NO. 6 as follows:

10 5' GAYGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGA
CCACCACGATGATCACCATGACGACCATGATGATGATGATGAAACAATGCACT
ATGCCCAGTGTGAAATGGAACCAAACCCTCATATGGCTAGCAGCCTTCACCA
CCATGTCCATGGCAGCATAGAGTTGTACAGAAGGGTCATGGAGCTGTTTAT
CTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA
15 TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTA
TTGGCGAACTGTACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCT
CGGTGACCTGGTTGACGATGATAGGGGCGTGTTAATGAAGTTCATCATTATG
CTTGTTGGACATTGATGGTACAGCACCAACACCGAAGCTCTCATTGGACA
CTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCAGCCAGTA
20 GAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGC
TGCTCTACATCACGAGCTAGAGGAAGATAAACTGAGCATTATGCCCATTGTG
ACGTAAGATCTAATACACACCAACCAAAGGCTCTTCATCATCATGTCCACGGA
ACCATCGATTTCAAACAAGTTGGTTATGGTGACCTTGAAGTGTCTACCATTTA
GAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGACGTACAGAT
25 CTACGCCAACGGTGACCTGACCAGTGAGTTGGGTGATCTAGGAGATATTCACGA
GATCCTCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGA
TGATGACCATGGCGTTGTCAATGAAAGCCACAGATATTCCTGGATCAATATCT
TCGGTGATGACAGTGTCTGGGACGTTCTATTGCCATTACCAAAGAGACCAT
CTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACGTGGACAGAGCCA
30 TCCAGAAATTGTTACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTAC
TGGTTTACATCACCATGTCTCTGGTTCTATAACATTCTGAACAGACCCCTGGAG
GATCAACACATATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTG
TCACATCATCGTCATGGTGTGCAGCTCCATGAATGGGGAGATATGTCCCATG
GCTGTCACTCCTTAGGCAGAAATGTACCATGGTCATGATGATGCTCATGACCCC
35 AAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATCGTTCA

5 A further polynucleotide has the sequence of SEQ ID NO. 8 as follows:

5'GAYGGGGAGCAGTGTAAACGATGGGCAGAACAAAGATGACCACCATGACGA
CCACCACGATGATCACCATGACGACCATGATGATGATGATGAAACAATGCACT
ATGCCCAGTGTGAAATGGAACCAAACCCTCATATGGCTAGCAGCCTTCACCA
10 CCATGTCCATGGCAGCATAGAGTTGTCACAGAAGGGTCATGGAGCTGTTTAT
CTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA
TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTA
TTGGCGAACTGTACAATGCTCACCAGAAAAACATGCTGACCCTGGTGACCT
CGGTGACCTGGTTGACGATGATAGGGGGCGTGGTTAATGAAGTTCATCATTATG
15 CTTGGTTGGACATTGATGGTACAGCACCAAAACACCGAAGCTCTCATTGGACA
CTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCCAGCCAGTA
GAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGC
TGCTCTACATCACGAGCTAGAGGAAGATAAACTGAGCATTATGCCCATTTGTG
ACGTAAGATCTAATACACACCAACCAAAGGCTCTTCATCATCATGTCCACGGA
20 ACCATCGATTTCAAACAAGTTGGTTATGGTGACCTTGAAGTGTCTACCATTTA
GAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGACGTACAGAT
CTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATAT
GATCCTCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGA
TGATGACCATGGCGTTGTCAATGAAAGCCACAGATATTCCTGGATCAATATCT
25 TCGGTGATGACAGTGTCTCTGGGACGTTCTATTGCCATTACCAAAGAGACCAT
CTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACGTGGACAGAGCCA
TCCAGAAATTGTTACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTAC
TGGTTTACATCACCATGTCTCTGGTTCTATAACATTGGAACAGACCCCTGGAG
GATCAACACATATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTG
30 TCACATCATCGTCATGGTGTGCAGCTCCATGAATGGGGAGATATGTCCCATG
GCTGTCACTCCTTAGGCAGAATGTACCATGGTCATGATGATGCTCATGACCCC
AAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATCGTTCA
TTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCT
TGTGATTATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGT
35 GTTATAGGACGGGCATGAATAACCTCACTAGAGTGACTTTGTCTAACATGACA

ATTAACAATTGTATAACTTCGCTAAAAAATAAAACAATGACACAATGNAAAAAA
 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA3'

with TGA being the opal stop codon and AATAAA the polyadenylation signal.

5

10 Variants or homologues of the above polynucleotide sequences also form part of the present invention. Polynucleotide sequences may be aligned, and percentage of identical nucleotides in a specified region may be determined against another sequence, using computer algorithms that are publicly available. Two exemplary algorithms for aligning and identifying the similarity of polynucleotide sequences are the BLASTN and FASTA algorithms. The BLASTN software is available on the NCBI anonymous FTP server (ftp://ncbi.nlm.nih.gov) under /blast/executables/. The BLASTN algorithm version 2.0.4 [Feb-24-1998], set to the default parameters described in the documentation and distributed with the algorithm, is preferred for use in the determination of variants according to the present invention. The use of the BLAST family of algorithms, including BLASTN, is described at NCBI's website at URL <http://www.ncbi.nlm.nih.gov/BLAST/newblast.html> and in the publication of Altschul, Stephen F, *et al* (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402. The computer algorithm FASTA is available on the Internet at the ftp site ftp://ftp.virginia.edu/pub/fasta/. Version 2.0u4, February 1996, set to the default parameters described in the documentation and distributed with the algorithm, is preferred for use in the determination of variants according to the present invention. The use of the FASTA algorithm is described in the W R Pearson and D.J. Lipman, 25 "Improved Tools for Biological Sequence Analysis," *Proc. Natl. Acad. Sci. USA* 85:2444-2448 (1988) and W.R. Pearson, "Rapid and Sensitive Sequence Comparison with FASTP and FASTA," *Methods in Enzymology* 183:63-98 (1990).

All sequences identified as above qualify as "variants" as that term is used herein.

30

Variant polynucleotide sequences will generally hybridize to the recited polynucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65°C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 35 1X SSC, 0.1% SDS at 65°C and two washes of 30 minutes each in 0.2X SSC, 0.1%

SDS at 65°C. Such hybridizable sequences include those which code for the equivalent protein from sources (such as shellfish) other than *P. canaliculus*.

While the above synthetic or recombinant approaches can be taken to produce the protein of the invention, it is however practicable (and indeed presently preferred) to obtain the protein by isolation from *P. canaliculus*. This reflects the applicants' finding that the protein is the dominant protein of the haemolymph of *P. canaliculus* and also that the protein is self-aggregating. It can therefore be isolated in commercially significant quantities direct from the mussel itself. For example, approximately 2 mg of the protein can be obtained per ml of haemolymph.

Once obtained, the protein is readily purified if desired. This will generally involve centrifugation in which the self-aggregating nature of the protein is important. Other approaches to purification (eg. chromatography) can however also be followed.

Furthermore, if viewed as desirable, additional purification steps can be employed using approaches which are standard in this art. These approaches are fully able to deliver a highly pure preparation of the protein.

Once obtained, the protein and/or its active fragments can be formulated into a composition. The composition can be, for example, a therapeutic composition for application as a pharmaceutical, or can be a health or dietary supplement. Again, standard approaches can be taken in formulating such compositions.

Still further, the composition can be a food in which the protein and/or its active fragments are included. This can occur by adding the protein to a pre-prepared foodstuff, or incorporating the protein into a step of the manufacturing process for the food.

The invention will now be described more fully in the following experimental section which is provided for illustrative purposes only.

EXPERIMENTAL

Section 1

5 **A. Materials and Methods**

10 **A.1 Shellfish:** *Perna canaliculus* (the New Zealand green-lipped mussel; the Greenshell™ mussel) were obtained at retail supermarket outlets or from mussel farmers directly; other shellfish species were obtained from retail outlets except for the blue mussel *Mytilus edulis* which was supplied by Sanford's Fisheries (Havelock, New Zealand).

15 **A.2 Extracts:** Mussel extracts were prepared by homogenising whole, shucked mussels (up to 120 mm length) in a commercial food processor with the addition of 0.02 M sodium phosphate buffer, pH 7.2. Dichloromethane (1/2 volume) was mixed with the aqueous extract, centrifuged at low speed (6000 rpm, GSA rotor, Sorvall RC-5B centrifuge at 4 °C). Polyethylene glycol (PEG) (MW 6000) was added to the aqueous phase to a final concentration of 10% (w/v) and NaCl to 0.5 M and stirred at 4-6 °C overnight. Following low speed
20 centrifugation the PEG-precipitate was resuspended in approximately 1/10 volume of sodium phosphate buffer. After another cycle of low-speed centrifugation the supernatant was centrifuged at high speed (50,000 rpm in a Beckman 60Ti rotor at 4 °C for 60-80 minutes). The resultant pellet was resuspended in a small volume of phosphate buffer and clarified by low
25 speed centrifugation.

30 **A.3 Polyacrylamide gel electrophoresis:** 12% polyacrylamide gels (8 x10 cm; 1 mm thick) were cast using a prepared stock solution according to the manufacturer's instructions (40% acrylamide/bis solution 37.5:1, Bio-Rad, USA); commercially available 12% gels (Bio-Rad, USA) were also used. Samples (10 µl) were applied to lanes and the gels run at 160 V using a standard Tris/Glycine/SDS buffer (Bio-Rad, catalogue 161-0732) until the bromphenol blue marker reached the bottom of the gel. Gels were stained with BM Fast Stain Coomassie® (Boehringer Mannheim, Germany) and
35 destained as per the manufacturer's instructions.

- 5 **A.4 Glycosylation t st:** Samples were treated with N-glycosidase F (PNGase F from *Flavobacterium meningosepticum*; Boehringer Mannheim Biochemica, Germany) according to the manufacturer's directions. Treated and untreated samples were run in a standard 12% polyacrylamide gel.
- 10 **A.5 Isopycnic gradients:** CsCl (Boehringer Mannheim, Germany) solutions were prepared in 0.1 M sodium phosphate buffer, pH 7.2 and filtered through a 0.22 µm membrane (Acrodisc, Gelman Sciences, USA) to clarify. Two step gradients (1.25 g/cc top layer containing the sample and 1.45 g/cc bottom layer) were prepared as described by Scotti (1985) and centrifuged for approximately 17 hours at 20 °C in a Beckman 70Ti rotor at 30,000 rpm. The resultant gradient was fractionated by inserting a 100 µl glass capillary tube into the gradient and slowly pumping out the contents. UV absorbance was monitored by passing through a Uvicord spectrophotometer (LKB Produkter, Sweden). Fractions were collected and the refractive indices measured using an Abbé refractometer (Bellingham and Stanley, UK) and the density estimated using regression equations according to the method of Scotti (1985).
- 15 **A.6 Porous glass chromatography:** Controlled pore glass (CPG 240-80, Sigma Chemical Co., USA) was treated according to the suppliers directions. A 1 cm x 100 cm column (Bio-Rad, USA) was prepared. Samples (1-2 ml) were loaded onto the column and eluted with 0.1 M sodium phosphate buffer, pH 7.2, through a Uvicord spectrophotometer, fractions being collected at regular intervals.
- 20 **A.7 Estimation of protein concentration:** Concentrations were estimated using a bovine serum albumin standard (Blot Qualified BSA, Promega, USA) by UV absorption according to the method of Layne (1957) using the equation: mg/ml protein = $1.55 \cdot A_{280} - 0.76 \cdot A_{260}$. Alternatively, concentration was estimated by the Bradford reaction using reagent supplied by Bio-Rad (USA) at a wavelength of 620 nm..
- 25
- 30

A.8 High performance liquid chromatography: Reversed-phase HPLC was performed on an HP 1050 Ti-series HPLC (Hewlett Packard, USA) fitted with an analytical 300 Å Vydac C-18 column, 25 cm x 4.6 mm i.d.. The 10 µl sample in water was eluted with a 0-100% acetonitrile in water (v/v) gradient over 60 min and the absorption at 218 and 280 nm was recorded.

B. Results

A light-scattering band was seen after centrifugation of extracts of whole Greenshell™ mussels in CsCl gradients (**Figures 1a and 1b**). The density of this band was estimated at 1.368 g/cc. A minor band was sometimes observed at approximately 1.390 g/cc. If rebanded in CsCl the 1.390 band yielded two bands - one at 1.390 g/cc and a second at 1.368 g/cc. SDS-PAGE analysis of fractions of either density gave similar polypeptide profiles with a single major band. The molecular weight of the protein by PAGE was estimated as 75,000 (75 kDa) (**Figure 1c**). Several minor bands of higher molecular weight and an additional minor band of 45 kDa were also seen. The main band (called pernin) at 75 kDa was always at great excess compared to the minor bands. When material from the light-scattering material from CsCl gradients were examined by electron microscopy, particles resembling those of "empty" small RNA viruses were seen (**Figure 2**). However a UV wavelength scan (data not shown) indicated that little, if any, nucleic acid was present and that the particles were mainly composed of protein. HPLC showed the CsCl band to be composed almost solely of a single species of protein (**Figure 3**). Since HPLC indicated a high degree of purity, the higher molecular weight polypeptides are presumed to be multimers of pernin. It is likely that the minor, lower molecular weight band is degraded pernin.

Chromatography, on a CPG 240-80 column, of semi-purified extracts, or of material banded in CsCl, showed that the majority of pernin was eluted in the exclusion volume using low molarity phosphate or Tris buffer as the eluent. In contrast, a protein of similar size, bovine serum albumin (68 kDa), was included in the column matrix. It appears, therefore, that pernin does aggregate into large, particle-like structures under certain conditions as suspected from the particles seen in **Figure 2**. HPLC confirmed that pernin from *P. canaliculus* obtained by CPG chromatography was highly purified. Aggregating protein species were also detected in extracts of

other shellfish: the blue mussel *Mytilus edulis*, the oyster *Crassostrea gigas*, and New Zealand pipis *Paphies australis* but not in scallops *Pecten novaezealandiae*. These polypeptides were lower in molecular weight than pernin (**Figure 4a**). The pernin from *P. canaliculus* is N-glycosylated as shown by a reduction in molecular weight when treated with endoglycosidase-F before PAGE (**Figure 4b**).

The yield of pernin from whole mussel extractions averaged about 200 µg/mussel. Improved yields of pernin were obtained by extracting haemolymph directly from live *P. canaliculus*. A small notch was made in the shell using a triangular file and a 30 gauge needle inserted into the posterior adductor muscle. From 1 to 5 ml of haemolymph can be withdrawn easily. The haemolymph was spun at low speed (≈1000 g) to remove haemocytes and the resulting supernatant processed by ultracentrifugation, for example at 250,000 g for 40 minutes, followed by either CPG chromatography eluting with 0.1 M sodium phosphate buffer, pH 7.2, or isopycnic banding in CsCl in phosphate buffer. The pernin obtained in this way appeared no different than that purified from whole mussels and had the advantage of a 30-fold average increase in yield from each mussel. Haemolymph contained around 2 mg/ml (average ≈5-6 mg/mussel) of pernin which is by far the most predominant polypeptide species (**Figure 5a**). The time to purify pernin was reduced from about 5 days to 1 day.

Microsequencing of the N-terminal region and internal fragments generated by chemical and enzymatic cleavage from purified pernin was performed and generated the following sequences of cleavage fragments:

- 25
- (a) DGEQCNDGQN
 - (b) QGGHEVESERVACCVIGRA
 - (c) GQSHPEIVH
 - (d) YHGHDDA
 - 30 (e) VVNEVVHH.

These sequences code for amino acids as follows:

CODE:

	A	alanine
	C	cystine
	D	aspartic acid
5	E	glutamic acid
	F	phenylalanine
	G	glycine
	H	histidine
	I	isoleucine
10	K	lysine
	L	leucine
	M	methionine
	N	asparagine
	P	proline
15	Q	glutamine
	R	arginine
	S	serine
	T	threonine
	V	valine
20	W	tryptophan
	Y	tyrosine

The sequence data was then compared with amino acid sequences in searchable computer data bases. Some sequences were found to be of particular interest:

25 The sequence data was then compared with amino acid sequences in searchable computer data bases. Some sequences were found to be of particular interest:

25 a) a 10 amino acid residue sequence from the N-terminus of pernin (sequence (a) above) showed only homology with an 8 base anti-thrombin protein sequence from terrestrial leeches (data from US Patent 5,455,181 Oct 3, 1995: sequence 10).

30

<i>Perna canaliculus</i> pernin	2	GEQCNDGQ	9
matching amino acids		G+ CNDGQ	
leech anti-thrombin	5	GQSCNDGQ	12

35

identities: 6/8 (75%) positives: 7/8 (87%);
 "+" indicates an equivalent amino acid;
 the bolded numerals indicate amino acid position

b) An internal cleavage product (sequence (b) above) was shown to have homology to the Cu-Zn class of proteins known as "SODs" (superoxide dismutases).

Each of fragments (a) to (e) are part of the larger pernin amino acid sequence:

5

1	DGEQCNDGQN	KDDHHDDHHD	DHHDDHDDDD	ETMHYAQCEM	EPNPHMASSL
5	HHHVHGSIEL	SQKGHGAVYL	ELHLVGFNTS	EDHDDHHHGL	HLHMLGDMSA
0	GCDSIGELYN	AHPEKHADPG	DLGDLVDDDR	GVVNEVHHYA	WLDIDGTAPN
5	TEALIGHSMT	ILQGSHTDAD	TPASRIACCV	IGHGKARPET	AAALHHELEE
20	DKTEHYAHCD	VRSNTHQPKA	LHHHVHGTID	FKQVGYGDL	VSYHLEGFNV
25	SDDHKDHLHD	VQIYANGDLT	SGCDNLGAKY	DPHEDYHSEL	GDLGDIHDDD
30	HGVVNESHRY	SWINIFGDDS	VLGRSIAIHQ	RDHLHKSAKI	ACCVIGRGQS
35	HPEIVHRAKC	VVRPNTSTG	LHHHVSGSIT	FEQTPGGSTH	MTADLKGFNV
40	SEDLSHRRHG	VQLHEWGDMS	HGCHSLGRMY	HGHDDAHDPK	RPGDLGDVID
45	DSHGIVHSTR	TFDHLNVEDL	NARSLVIMQG	GHEVESERVA	CCVIGRA

(Bold characters indicate directly sequenced fragments (a) to (e)).

10 Section 2

Anti-thrombin Activity

The possibility that pernin could function as an anti-thrombin agent was examined in a kinetic assay for thrombin inhibition.

Thrombin inhibition assay

Kinetic assays were done using an Accucolor™ Antithrombin III kit (catalogue no. CRS105, Sigma Diagnostics, USA) with the reagents prepared according to the supplier's directions. Standard plasma was supplied by Instrumentation Laboratories (Italy) and used at the recommended dilution of 1/41. Samples of purified mussel protein in water were diluted 9/10 by adding 10X Sigma sample buffer. Heparin was purchased from Instrumentation Laboratories. Thrombin activity was estimated colorimetrically at 405 nm using a chromogenic substrate (H-D-HHT-L-Ala-L-Arg-pNa.2AcOH, catalogue no. A 8058, Sigma, USA) and a Multiskan Biochromatic plate reader (Labsystems, Finland)

This verified that pernin had inhibitory activity. When a purified preparation of pernin was centrifuged through a 30,000 MW exclusion filter (**Figure 5a**), all the anti-thrombin activity was in the retentate and no detectable activity was present in

the filtrate (**Figure 5b**). The standard serum was diluted 1/41 as recommended for this assay system; the pernin concentration was not determined directly but was in the 1 mg/ml range. From this kinetic data pernin inhibition was estimated to be about 50% of the level of human plasma (approximately 1 mg/ml pernin diluted 9/10 compared with the 1/41 plasma dilution in the standard ATIII assay system). Heparin, a co-factor required for ATIII inhibition of thrombin, was not required for inhibitory action by pernin.

Metal Binding Activity

10

Hi Trap® Chelating affinity columns (Amersham Pharmacia Biotech, 1ml size) were prepared according to the manufacturer's instructions. The columns were then charged with either 0.1M cupric chloride or zinc chloride before equilibrating in a buffer (0.050M sodium phosphate and 0.5M sodium chloride containing 0.5mM imidazole, pH 7.0). Protein samples purified using CsCl centrifugation were suspended in this buffer and applied to the column using a chromatographic system (Econo System, Bio-Rad Laboratories, USA). Following washing of the column for 5 mins with buffer during which no protein appeared in the eluate, a linear gradient over 20 min at 1 ml/min was used to develop the column using buffer with the imidazole concentration at 100mM from 0-100%. The protein eluted into the gradient being retained longer on the copper chelation column than the zinc. The absorption of the eluate was monitored at 254nm.

Divalent metal ion content of the CsCl purified protein was determined by dissolving the protein in water at 10 mg/ml and analysing metal content by both atomic absorption and plasma emission spectrometry by comparison with a water blank. There was no significant divalent cation content in the protein purified by this method. However, purification by other methods not employing chaotropic agents like CsCl, the high content of histidine coupled with acidic amino acid residues and the likely origin of this protein from a SOD precursor, points to pernin having endogenous metal ions as part of its native structure.

Section 3**Gen Sequencing Method**

5 ~~A suite of non-specific primers called pUZ5 was synthesised by Gibco-BRL for the initial sequencing based on the N-terminal sequence of pernin. The general formula was:~~

~~GAY GGN GAR CAR TGY AAY GAY GGN CAR AA~~

10

~~Where Y represents a pyrimidine base, R represents a purine base and N represents any one of the four nucleotide bases. Sequencing was done, initially using pUZ5 and an oligo-dT based "bottom stand" primer from PCR amplified cDNA. Sequencing was done by dye-termination cycle sequencing using "BigDye" prism technology (Applied Biosystems Incorporated, USA) according to their instructions. Products were resolved on an ABI 377 automated sequencer. Following the initial sequencing of approximately 500 base pairs pernin-specific primers were constructed and used to complete the sequencing of the pernin gene.~~

20 This provided the following:

GAYGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGACCACCACGATGATCA
CCATGACGACCATGATGATGATGATGAAACAATGCACTATGCCAGTGTGAAATGGAACCAAACC
CTCATATGGCTAGCAGCCTTCACCACCATGTCCATGGCAGCATAGAGTTGTACAGAAGGGTCAT
25 GGAGCTGTTTATCTAGAACTTCATCTTGTCTGGATTCAACACAAGTGAAGACCATGACGACCACCA
TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTATTGGCGAACTGT
ACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCTCGGTGACCTGGTTGACGATGATAGG
GGCGTGTTAATGAAGTTCATCATTATGCTTGGTTGGACATTGATGGTACAGCACCACCAACACCGA
AGCTCTCATTTGGACACTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCAGCCA
30 GTAGAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCAGAAACAGCAGCTGCTCTACAT
CACGAGCTAGAGGAAGATAAACTGAGCATTATGCCCATTTGTGACGTAAGATCTAATACACACCA
ACCAAAGGCTCTTCATCATCATGTCCACGGAACCATCGATTTCAAACAAGTTGGTTATGGTGACC
TTGAAGTGTCTTACCATTTAGAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGAC
GTACAGATCTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATATGATCC
35 TCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGATGATGACCATGGCGTTG
TCAATGAAAGCCACAGATATTCTGGATCAATATCTTCGGTGATGACAGTGTCTGGGACGTTCT
ATTGCCATTACCAAAGAGACCATCTTCATAAAAGTGCCAAAATTGCCTGTTGTGTATAGGACG
TGGACAGAGCCATCCAGAAATTGTTACAGAGCTAAATGTGTTGTGACACCTAATACAGAATCTA
CTGGTTTACATCACCATGTCTCTGGTTCTATAACATTGGAACAGACCCCTGGAGGATCAACACAT

ATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTGTCACATCATCGTCATGGTGTGCA
 GCTCCATGAATGGGGAGATATGTCCCATGGCTGTCACTCCTTAGGCAGAATGTACCATGGTCATG
 ATGATGCTCATGACCCCAAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATC
 GTTCATTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCTTGTGAT
 5 TATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGTGTTATAGGACGGGCATGAA
 TAACCTCACTAGAGTGACTTTGTCTAACATGACAATTAACAATTGTATAACTTCGCTAAAAAATA
 AAACAATGACACAATGNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA.

Discussion

10

The present invention is a novel protein obtainable from *Perna canaliculus*, the New Zealand green-lipped (Greenshell™) mussel. The protein appears to be able to self-aggregate in structures resembling small virus like particles (VLPs) approximately 25 nm in diameter but lacking any nucleic acid. The protein was
 15 found in extracts of whole mussels and appears to be the predominant protein in haemolymph. The molecular weight of the protein was estimated to be 75 kDa by PAGE and inferred to be 55 kDa from its polynucleotide encoding sequence but, because of its ability to aggregate, the protein can be sedimented by ultracentrifugation in a short time (e.g. 40 minutes at 250,000 *g*) whereas the
 20 monomeric protein would not. Each ml of haemolymph yields, on the average, about 2 mg of pernin. Haemolymph is easily obtained by withdrawing fluid from the posterior adductor muscle of the shellfish which can yield up to 5 ml without obvious harm; it is not necessary to kill the mussel. The haemolymph obtained not only contains high levels of pernin but is quite free of contaminating materials,
 25 particularly compared with whole mussel extracts, so purification of pernin is simple. For highly pure preparations of pernin, ultracentrifugation is followed by isopycnic banding in a suitable density gradient medium such as CsCl.

The sequence of the N-terminus of pernin suggested that the protein might have
 30 anti-thrombin activity. This was demonstrated in kinetic assays on purified pernin. Since thrombin is a serine protease, pernin also acts as a serine protease inhibitor.

Comparison of the sequences obtained from several cleavage fragments against amino acid sequences in a computer database suggest that in addition to the anti-
 35 thrombin activity of pernin, the protein also possesses other activities. One of these is the ability to bind divalent cations such as Zn²⁺ and Cu²⁺.

INDUSTRIAL APPLICATION

The preferred protein of the invention, pernin, has a number of utilities.

5

Because of its anti-thrombin activity pernin is potentially useful as an anti-coagulant agent. Thrombin normally acts as a protease which converts fibrinogen in the blood to fibrin. Blood coagulation is counteracted by inhibitors, normally anti-thrombin III (ATIII); pernin has also been shown to inhibit thrombin activity in an
10 ATIII assay system. In contrast to ATIII, whose action is accelerated by the presence of heparin (a sulphated mucopolysaccharide) pernin does not require heparin as a co-factor.

The pernin protein from *P. canaliculus* thus has value as a pharmaceutical. Since it
15 is active as an anticoagulant in its native state it may also be useful as a natural therapeutic agent or health supplement. It is readily obtained as a natural product in high concentrations from mussel haemolymph. To obtain a highly pure preparation it is necessary only to remove haemocytes by centrifugation (or any other suitable method) followed by either ultracentrifugation (since pernin forms
20 aggregates which readily sediment) and resuspension, isopycnic banding in a suitable medium such as CsCl, exclusion filtration through a suitable membrane which retains pernin, or chromatography through a medium such as controlled pore glass of suitable porosity. The result is a highly pure preparation of pernin.

25 The mussel *P. canaliculus* produces large amounts of the protein naturally, with little cost or effort involved in production, processing or purification.

A further utility of the protein arises from the fact that pernin can be stripped of divalent cations (for example by CsCl isopycnic banding, or pH variation). This
30 allows for the addition of divalent cations of choice (such as Mg^{++} , Cd^{++} , Zn^{++} or Ca^{++}) to the metal stripped pernin. Such a protein, with a modified and pre-selected divalent cation loading, has application in the food and nutraceutical industries.

The ability to bind divalent metal cations also gives rise to applications of the
35 protein in bioremediation and/or cation recovery processes. The divalent cations

can be present as contaminants or pollutants in, for example, a solution, and the solution passed by a substrate to which the protein is bound so that the cations are extracted.

- 5 Yet a further utility arises from the fact that the protein is "self-aggregating", and can form into structures resembling empty virus-like particles of approximately 25 nm in diameter. These empty virus-like particles are able to sequester other molecules inside them, with the consequent ability to function as delivery vehicles for those other molecules. Examples of molecules able to be delivered in this
10 manner include pharmaceutically active compounds.

Those persons skilled in the art will understand that the above description is provided by way of illustration only and that the invention is limited only by the appended claims.

15

REFERENCES

Layne, E. (1957). Spectrophotometric and turbidometric methods for measuring proteins, *Methods in Enzymology* **III**, 447.

20

Scotti, P.D. (1985). The estimation of virus density in isopycnic cesium chloride gradients. *Journal of Virological Methods* **12**, 149.